

WHAT IS CLAIMED IS:

- Sub B1
- 3 1. A method for parallelizing a computer application program based on a script of a script-driven software tool, comprising automatically analyzing the script and producing a parallel-computation specification based on such analysis, where such parallel
- 4 computation specification provides functional equivalence to the script when executed by
- 5 a parallel runtime system.
- 1 2. A method for parallelizing a computer application program based on a script of a script-driven software tool, comprising automatically analyzing the script and producing a
- 2 parallel computation specification plus a script fragment set based on such analysis,
- 3 where such parallel computation specification and script fragment set provides functional
- 4 equivalence to the script when executed by a parallel runtime system.
- 5
- SUB A1
- 2 3. The method of claims 1 or 2, wherein automatically analyzing the script includes:
- 3 (a) parsing the script into statements;
- 4 (b) constructing a serial dataflow graph from the parsed statements; and
- 1 (c) construction a parallel dataflow graph from the serial dataflow graph.
- 2 4. The method of claim 3, wherein constructing the serial dataflow graph includes:
- 3 (a) constructing a serial dataset table of datasets used by the script;
- 4 (b) constructing a serial processing step table of statements performed by the script; and
- 5 (c) constructing a serial dataset access table indicating datasets in the dataset table used
- 1 by statements in the processing step table.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
1  
2  
1  
2  
3  
1  
1  
2  
3  
4  
1

5. The method of claim 3, wherein constructing the parallel dataflow graph includes:
  - (a) constructing a parallel dataset table of datasets based on the serial dataset table;
  - (b) constructing a parallel processing step table of statements based on the serial processing step table;
  - (c) constructing a dataset access table based on the serial dataset access table; and
  - (d) determining, for each processing step identified in the parallel processing step table, if a corresponding pre-defined parallelization rewrite rule exists for such processing step, and if so, then applying the corresponding pre-defined parallelization rewrite rule to redefine associated entries in the parallel dataset table, the parallel processing step table, and the dataset access table as parallel processing entries; and if not, then defining such associated entries as serial processing entries.
6. The method of claim 5, further including resolving any existing partitioning conflicts in the constructed parallel dataflow graph.
7. The method of claim 5, wherein at least one pre-defined parallelization rewrite rule is an algorithm selected from the group comprising simple partitioning, key-based partitioning, local-global division, external parallelism algorithm, and statement decomposition.
8. The method of claims 1 or 2, wherein the script-driven software tool is SAS®.
9. The method of claims 1 or 2, wherein producing the parallel computation specification includes applying at least one pre-defined parallelization rewrite algorithm selected from the group comprising simple partitioning, key-based partitioning, local-global division, external parallelism algorithm, and statement decomposition.

Sub A27  
Page 10 of 20

1 10. A computer program, residing on a computer-readable medium, for parallelizing a  
2 computer application program based on a script of a script-driven software tool, the  
3 computer program comprising instructions for causing a computer to automatically  
4 analyze the script and produce a parallel computation specification based on such  
5 analysis, where such parallel computation specification provides functional equivalence  
6 to the script when executed by a parallel runtime system.

1 11. A computer program, residing on a computer-readable medium, for parallelizing a  
2 computer application program based on a script of a script-driven software tool, the  
3 computer program comprising instructions for causing a computer to automatically  
4 analyze the script and produce a parallel computation specification plus a script fragment  
5 set based on such analysis, where such parallel computation specification and script  
6 fragment set provides functional equivalence to the script when executed by a parallel  
7 runtime system.

1 12. The computer program of claims 10 or 11, wherein automatically analyzing the script  
2 includes:

- 3 (a) parsing the script into statements;  
4 (b) constructing a serial dataflow graph from the parsed statements; and  
5 (c) construction a parallel dataflow graph from the serial dataflow graph.

1 13. The computer program of claim 12, wherein constructing the serial dataflow graph  
2 includes:

- 3 (a) constructing a serial dataset table of datasets used by the script;  
4 (b) constructing a serial processing step table of statements performed by the script; and  
5 (c) constructing a serial dataset access table indicating datasets in the dataset table used  
6 by statements in the processing step table.

1 14. The computer program of claim 12, wherein constructing the parallel dataflow graph  
2 includes:

3 (a) constructing a parallel dataset table of datasets based on the serial dataset table;  
4 (b) constructing a parallel processing step table of statements based on the serial  
5 processing step table;  
6 (c) constructing a dataset access table based on the serial dataset access table; and  
7 (d) determining, for each processing step identified in the parallel processing step table,  
8 if a corresponding pre-defined parallelization rewrite rule exists for such processing  
9 step, and if so, then applying the corresponding pre-defined parallelization rewrite  
10 rule to redefine associated entries in the parallel dataset table, the parallel processing  
11 step table, and the dataset access table as parallel processing entries; and if not, then  
12 defining such associated entries as serial processing entries.

1 15. The computer program of claim 14, further including resolving any existing partitioning  
2 conflicts in the constructed parallel dataflow graph.

1 16. The computer program of claim 14, wherein at least one pre-defined parallelization  
2 rewrite rule is an algorithm selected from the group comprising simple partitioning, key-  
3 based partitioning, local-global division, external parallelism algorithm, and statement  
4 decomposition.

1 17. The computer program of claims 10 or 11, wherein the script-driven software tool is  
2 SAS®.

1 18. The computer program of claims 10 or 11, wherein producing the parallel computation  
2 specification includes applying at least one pre-defined parallelization rewrite algorithm  
3 selected from the group comprising simple partitioning, key-based partitioning, local-  
4 global division, external parallelism algorithm, and statement decomposition.

-47-